



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/629,175	07/31/2000	Ophir Frieder	7519-164345	4562
26694	7590	01/13/2004	EXAMINER	
VENABLE, BAETJER, HOWARD AND CIVILETTI, LLP P.O. BOX 34385 WASHINGTON, DC 20043-9998			LE, UYEN T	
		ART UNIT	PAPER NUMBER	
		2171	12	

DATE MAILED: 01/13/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/629,175	FRIEDER ET AL.	
	Examiner Uyen T. Le	Art Unit 2171	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) Responsive to communication(s) filed on 03 December 2003.

2a) This action is **FINAL**. 2b) This action is non-final.

3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) Claim(s) 1-42 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5) Claim(s) _____ is/are allowed.

6) Claim(s) 1-42 is/are rejected.

7) Claim(s) _____ is/are objected to.

8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) The specification is objected to by the Examiner.

10) The drawing(s) filed on _____ is/are: a) accepted or b) objected to by the Examiner.

 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. §§ 119 and 120

12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

 a) All b) Some * c) None of:

 1. Certified copies of the priority documents have been received.

 2. Certified copies of the priority documents have been received in Application No. _____.

 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

 * See the attached detailed Office action for a list of the certified copies not received.

13) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application) since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.

 a) The translation of the foreign language provisional application has been received.

14) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121 since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.

Attachment(s)

1) Notice of References Cited (PTO-892) *Uyen*
 2) Notice of Draftsperson's Patent Drawing Review (PTO-948)
 3) Information Disclosure Statement(s) (PTO-1449) Paper No(s) 10.

4) Interview Summary (PTO-413) Paper No(s) _____ .
 5) Notice of Informal Patent Application (PTO-152)
 6) Other: _____ .

DETAILED ACTION

Response to Amendment

1. Applicant's amendment filed 3 December 2003 has been entered. Applicant points out to page 15, line 13 through page 16, line 20 as supporting the claimed "single tuple" and "single hash value" of a document. Therefore, the finality of the previous Office Action is withdrawn. However, due to prior art references submitted by the applicant, new grounds of rejection are presented in this Office Action.

2. Applicant's amendment to the specification to remove embedded link is acknowledged. However, applicant has not removed the new subject matter introduced by the amendment to the specification regarding Figure 2. Applicant is requested to remove the new matter introduced in the specification.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

3. Claims 34-42 are rejected under 35 U.S.C. 102(b) as being anticipated by Udi Manber "Finding similar files in a large file system", JSENIX, January 17-21, 1994 provided by the applicant.

Regarding claim 34, Manber discloses all the claimed subject matter including determining a hash value for a document, accessing a document storage structure

comprising a plurality of hash values, each hash value representing one of a plurality of documents, determining if the hash value is equivalent to another hash value in the document storage structure (see pages 4-5).

Claims 35, 37 correspond to a system to perform the method of claim 34, thus are rejected for the same reasons stated in claim 34 above.

Claim 36 corresponds to a computer program product to perform the method of claim 30, thus is rejected for the same reasons stated in claim 34 above.

Regarding claim 38, Manber discloses a method for detecting similar documents including comparing a document to a plurality of documents in a document collection using a hash algorithm and collection statistics to detect if the document is similar to any of the documents in the document collection (see pages 4-5).

Regarding claim 39, clearly the collection statistics has to pertain to the document collection since the statistics are used in clustering the collected documents.

Claims 40, 42 correspond to a system to perform the method of claim 38, thus are rejected for the same reasons stated in claim 38 above.

Claim 41 corresponds to a computer program product to perform the method of claim 38, thus is rejected for the same reasons stated in claim 38 above.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 1, 2, 5, 9, 12-19, 24, 25, 27-29 are rejected under 35 U.S.C. 103(a) as being unpatentable over Brin et al "Copy detection mechanisms for digital documents", ACM 1995, pages 398-409, in view of Haber et al (US 5,136,646) of record.

Regarding claim 1, Brin discloses a method for detecting similar documents (see the abstract). Brin discloses obtaining a document and filtering the document to obtain a filtered document (see page 400, left column, canonical form document). Brin teaches all the claimed concept of detecting similarity by comparing tuples. However, Brin breaks up a document into chunks then uses hashing to detect matching chunks. Brin clearly shows a tuple representing one chunk and an entry is stored in the hash table for every tuple (see page 400, right column, last paragraph). Claim 1, last paragraph merely reads on the fact that the method of Brin compares tuples to determine similar chunks. Haber shows hashing a document into a single hash value (see the abstract). Since a document hash value uniquely identifies a document, it would have been obvious to one of ordinary skill in the art to apply the principles taught by Brin to a whole document in order to detect document similarity instead of just portions of a document.

Regarding claim 2, Brin discloses parsing and filtering the document when Brin shows the document in canonical form (see page 400, left column, 3rd paragraph). Clearly the filtered document comprises a token stream of a plurality of tokens as claimed.

Regarding claim 5, Brin discloses determining the hash value for the filtered document by processing individually each retained token in the token stream when

Aiken show processing chunks and determining their hash values (see page 400, right column).

Regarding claim 9, although Brin and Haber do not explicitly disclose filtering by removing a duplicate of another token in the token stream, it would have been obvious to one of ordinary skill in the art to include such a feature in order to avoid processing redundant token, thus saving time and resources.

Regarding claim 12, Brin discloses removing formatting from the document (see page 400, left column).

Regarding claims 13, 14, clearly the method of Brin and Haber uses collection statistics pertaining to a plurality of documents for filtering the document since the document chunks are compared to a set of collected chunks to detect similarity (see page 400).

Regarding claims 15-18, although Brin and Haber do not explicitly show that the method uses specific hash algorithms as claimed, it is notoriously well known in the art to use different hash algorithms depending on users' requirements. Therefore, it would have been obvious to one of ordinary skill in the art to include all the claimed features while implementing the method of Aiken in order to suit users' needs.

Regarding claim 19, Brin discloses a hash table (see page 400, right column, last paragraph).

Regarding claim 24, Brin discloses inserting the tuple into the document storage structure (see page 400, right column).

Regarding claim 25, the hash table of Brin clearly comprises a plurality of bins of tuples as claimed and the step of determining if the tuple is clustered with another tuple clearly comprise determining if the tuple is co-located with another tuple at a bin of a hash table (see page 400, right column).

Claims 27, 29 correspond to a system to perform the method of claim 1, thus are rejected for the same reasons stated in claim 1 above.

Claim 28 corresponds to a computer program product to perform the method of claim 1, thus is rejected for the same reasons stated in claim 1 above.

5. Claims 3, 4, 6-8, 10, 11, 20-23, 26, 30-33 are rejected under 35 U.S.C. 103(a) as being unpatentable over Brin et al "Copy detection mechanisms for digital documents", ACM 1995, pages 398-409, in view of Haber et al (US 5,136,646) of record, in view of Aiken (US 6,240,409) of record.

Regarding claim 3, although Brin and Haber do not specifically show the claimed features, Aiken discloses retaining a token according to at least a token threshold when Aiken shows that common or frequent words are removed (see column 4, lines 38-53). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed features while implementing the method of Brin and Haber in order to filter the documents.

Claim 4 merely reads on the canonical document in the method of Brin (see page 400).

Regarding claim 6, although Brin and Haber do not specifically show the claimed features, Aiken discloses determining a score for each token in the token stream and comparing the score for each token to a first token threshold when Aiken shows that common or frequent words are removed. A threshold has to be present for the method of Aiken to determine common and frequent words (see column 4, lines 38-53). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed features while implementing the method of Brin and Haber in order to filter the documents.

Regarding claim 7, although Brin and Haber do not specifically show the claimed features, Aiken teaches the concept of selectively storing substrings (see column 6, lines 29-30). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed features while implementing the method of Brin and Haber in order to further filter the document and save memory.

Regarding claim 8, although Brin and Haber do not specifically show the claimed features, Aiken discloses filtering by removing from the token stream at least one token corresponding to a stop word (see column 4, lines 57-58, column 8, line 67- column 9, line 3). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed features while implementing the method of Brin and Haber in order to further filter the documents.

Regarding claim 10, although Brin and Haber do not specifically show the claimed features, Aiken discloses removing a token from a token stream based on collection statistics and at least one token threshold when Aiken shows that the method

remove words of “the” “and” , “this”, “is” (see column 4, lines 57-58, column 8, line 67- column 9, line 3). A token threshold has to be present for the method of Aiken to determine common and frequent words (see column 4, lines 38-53). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed features while implementing the method of Brin and Haber in order to further filter the documents.

Regarding claim 11, although Brin and Haber do not specifically show the claimed features, Aiken discloses removing a token from a token stream (see column 4, lines 57-58, column 8, line 67- column 9, line 3). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed features while implementing the method of Brin and Haber in order to further filter the documents.

Regarding claim 20, although Brin and Haber do not specifically show a tree, Aiken discloses that the document storage structure comprises a tree (see column 8, lines 30-38). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed features while implementing the method of Brin and Haber in order to represent a hierarchy of documents.

Regarding claims 21, 22, Aiken discloses that the tree comprises a binary tree (see column 8, lines 36-38). Although Aiken does not explicitly show that the binary tree is balanced, it would have been obvious to one of ordinary skill in the art to include such a feature in order to store data efficiently and to facilitate searching and localization.

Regarding claim 23, Brin discloses a hash table (see page 400, right column). Furthermore, Aiken discloses a hash table and at least one tree (see column 5, lines 33-40, column 8, lines 30-38). Therefore, it would have been obvious to one of ordinary

skill in the art to include the claimed features while implementing the method of Brin and Haber in order to represent documents in various forms.

Regarding claim 26, although Brin and Haber do not specifically show the claimed features, Aiken discloses a tree comprising a plurality of branches, each bucket of the tree comprising at least one tuple and wherein the step of determining if the tuple is clustered with another tuple clearly comprise determining if the tuple is co-located with another tuple in a bucket of the tree (see column 8, lines 31-54, Figure 4c). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed features while implementing the method of Brin and Haber in order to compare documents represented in a tree.

Claim 30 is a mere combination of claims 1-4, 26, thus is rejected for the same reasons stated in claims 1-4, 26 above.

Claims 31, 33 correspond respectively to a computer and system to perform the method of claim 30, thus are rejected for the same reasons stated in claim 30 above.

Claim 32 corresponds to a computer program product to perform the method of claim 30, thus is rejected for the same reasons stated in claim 30 above.

Conclusion

6. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not

mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

7. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Uyen T. Le whose telephone number is 703-305-4134. The examiner can normally be reached on M-F 7:00-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Safet Metjahic can be reached on 703-308-1436. The fax phone number for the organization where this application or proceeding is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-3900.


Uyen Le
Primary Examiner
AU 2171

12 January 2004